Building and Measuring a High Performance Network Architecture

for

CS 252

December 12, 2000

Bill Kramer

Contact: kramer@lbl.gov

510 486-7577

SCinet Team

Aaronsen Group, Argonne National Laboratory (DOE), Army Research Laboratory (DOD), Avici Systems, Caltech, Corp of Engineers Waterways Experimental Station (DOD), CISCO Systems, the Dallas Convention Center, the Dallas Convention and Visitor's Bureau, Extreme Networks, Foundry, GST Telecom, Internet-2 (NSF), Juniper Networks, Lawrence Berkeley National Laboratory (DOE), Lawrence Livermore National Laboratory (DOE), Marconi, MCI, MITRE Corporation, National Center Scientific Applications (NSF), Northeast Regional Data Center/ University of Florida, Nichols Research/CSC, Nortel Networks, Oak Ridge National Laboratory (DOE), Oregon State University, Pacific Northwest Laboratory (DOE), Qwest Communications, Sandia National Laboratory (DOE), Spirent Communications, Texas A&M University, University Corporation for Advanced Internet Development, University of Tennessee/Knoxville, the Very high performance Backbone Network Services - vBNS (NSF)

SCinet 2000 Goals

- Provide a stable, high quality network infrastructure for SC 2000 activities
  - support the basic infrastructure of the conference, the attendee's needs and the extensive education program.
- Provide reliable, high performance network for exhibitors
  - conference provides a showcase for the research and development activities of both the industry and research exhibitors.
- Provide experimental opportunities and demonstrations for the latest High Performance Networking technology
  - the opportunity to co-locate so much technology, expertise and so many varied applications seldom comes and a role of SCinet is to demonstrate the use of this technology in novel ways.
- Support and facilitate applications that make use of high performance networks it is one thing to build a large network, it is another to make it usable and still another to actually use it. SCinet works to accomplish all three objectives.

Four Separate Networks

- COMMODITY
- WIRELESS
- PRODUCTION
- EXPERIMENTAL

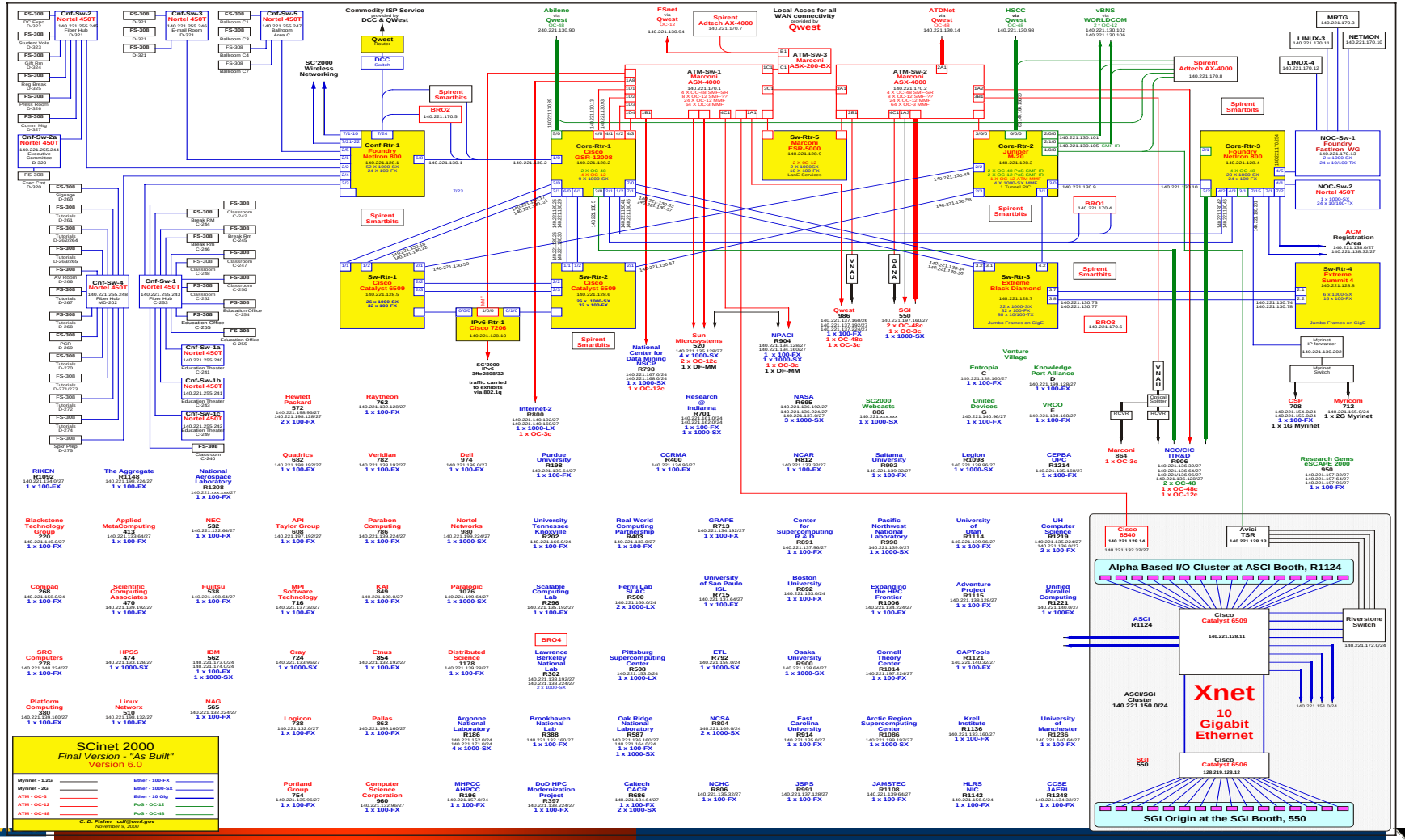
- SCinet is a full fledged ISP

Internal Connectivity

Type of Connection	Number of Connection	
	SC 99	SC 2000
OC-48c ATM	2	6
OC-48 PoS	1	5
OC-12c ATM	5	13
OC-12 PoS	0	2
OC-3c ATM	11	7
1,000 Mbps-LX	0	5
1,000 Mbps-SX	29	67
100 Mbps-FX	46	79
10 Mbps-FL	22	0

- Adding wireless, Commodity and other – 130 Gbps of total bandwidth

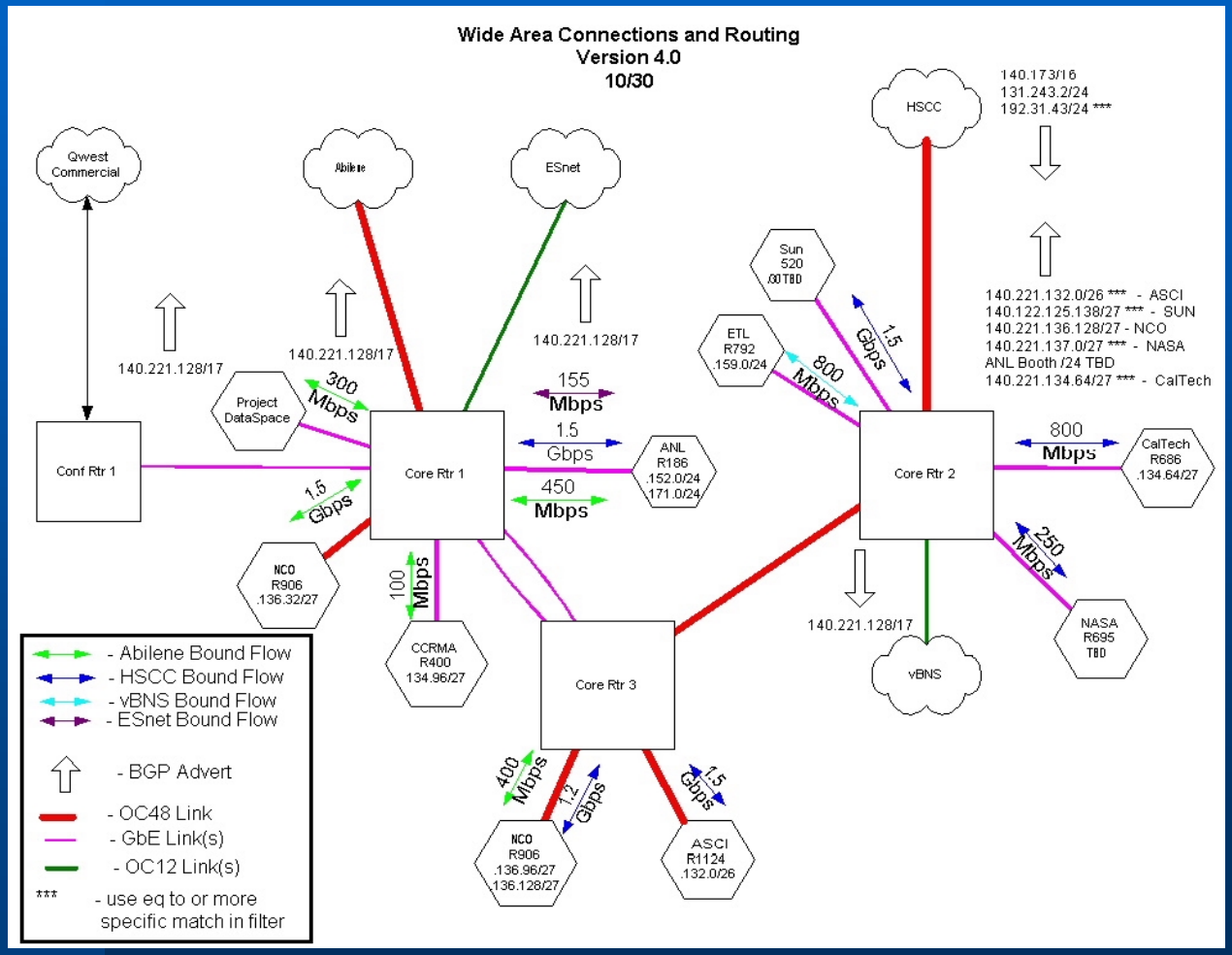
SCinet Design



External Connectivity

Network	Type	Maximum Speed
Abilene	OC-48 ATM	2.5 Gbps
ATDnet	OC-48 ATM	2.5 Gbps
HSCC	OC-48	1.5 Gbps
Esnet	OC 12 ATM	655 Mbps
vBNS	OC 12 ATM	655 Mbps
vBNS	OC 12 POS	655 Mbps
Commodity	ATM	12 Mbps
Total		8.477 Gbps

SCinet WAN Flow Diagram



SCinet Timeline

Event	Date	Time from Start
Arrival	10/30/00 9:00 AM	
First Fiber Lifted	10/30/00 4:00 PM	7.00
First Newspaper article	11/1/00 8:00 AM	
First Light to DCC	11/1/00 7:01 PM	58.02
First Power to NOC Racks	11/1/00 7:48 PM	58.80
First Light Extended to Electronics (GSR)	11/1/00 8:43 PM	59.72
Foundry NetIron on-line	11/1/00 8:56 PM	59.93
Cisco GSR, Foundry NI on-line	11/2/00 12:17 PM	75.28
Fore ASX40000 #1 on-line	11/2/00 1:32 PM	76.53
Fore ASX 4000 #2 on-line	11/2/00 1:35 PM	76.58
Juniper M20 on-line	11/2/00 3:20 PM	78.33
Extreme Black Diamond on-line	11/2/00 4:10 PM	79.17
Abilene Circuit up to Dallas POP	11/2/00 4:10 PM	79.17
Extreme Summit 4 on-line	11/2/00 4:15 PM	79.25
Marconi ESR 5000 on-line	11/2/00 4:20 PM	79.33
Cisco 7507 on-line	11/2/00 4:22 PM	79.37
SPIRENT gear (collectively) on-line	11/2/00 5:09 PM	80.15
Abilene Circuit Completed - First OC 48 for SC2000	11/2/00 5:14 PM	80.23
Abilene Peering up	11/2/00 5:25 PM	80.42
Cisco 6509 #1 on line	11/2/00 5:44 PM	80.73
Cisco 6509 # 2 on-line	11/2/00 5:45 PM	80.75
vBNS OC12 Packet over Sonnet up	11/2/00 6:40 PM	81.67
Best Power UPS on-line; All power changes complete	11/3/00 1:40 PM	100.67
HSCC OC-48 POS up	11/3/00 2:30 PM	101.50
Wireless Aps installed in Education area	11/3/00 5:00 PM	104.00
Address data purified and in the DB	11/4/00 7:15 AM	118.25
GPS Working for network monitoring	11/4/00 12:00 PM	123.00
Second vBNS OC-12 POS up, ATM Juniper, Marconi Up	11/4/00 12:45 PM	123.75
Completed Helpdesk Software	11/4/00 12:59 PM	123.98
Began accepting drop requests	11/4/00 1:00 PM	124.00
First Video Conference DCC to NSF at 30 frames a second	11/4/00 2:00 PM	125.00
ESnet up	11/5/00 1:35 PM	148.58
Bro 3 tap	11/5/00 3:00 PM	150.00
First TV report on local ABC affiliate	11/5/00 5:30 PM	152.50
260 Wireless Clients	11/6/00 11:00 AM	170.00
Bandwidth challenge 1.56 Gbps 1 second sample peak	11/7/00 10:15 PM	205.25
SCinet Production Network Shut down	11/9/00 4:00 PM	247.00
SCinet completely torn down and shipped	11/10/00 6:30 PM	273.50
Total Time from set up to tear down		11 Days, 12 Hours, 30 Seconds

What does it look like

Measurement and Testing

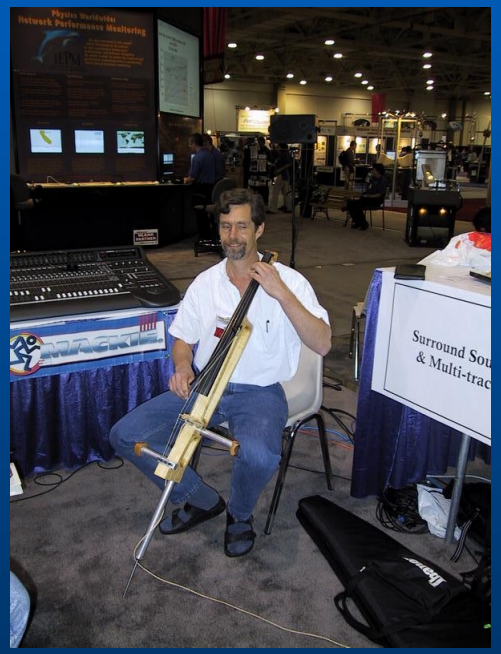
- Spirent Systems "Smart Bits" and Adtech technology to monitored and measured aspects of SCinet.
- The Indiana University Network Administration Suite (IUNAS) Internet-2 “Weathermap” technology for monitor wide area flows.
- Further measurement by Netflow software package
- The "Bro" package from LBNL used to monitor network traffic for intrusion.
- The SCinet team created custom software to measure other aspects of the network such as the wireless usage.
  - Spirent, Adtech and Bro used optical splitters to tap into the actual network connections at various points in the network.

Network Bandwidth Challenge

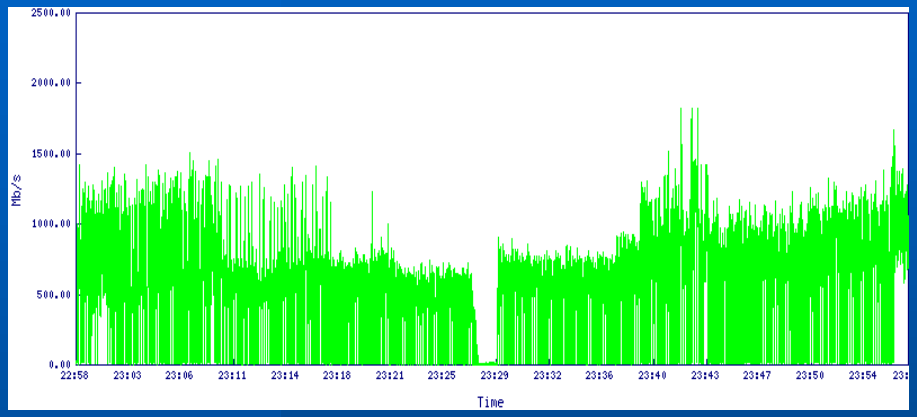
Applications specifically challenged to "use all the network bandwidth"

Examples:

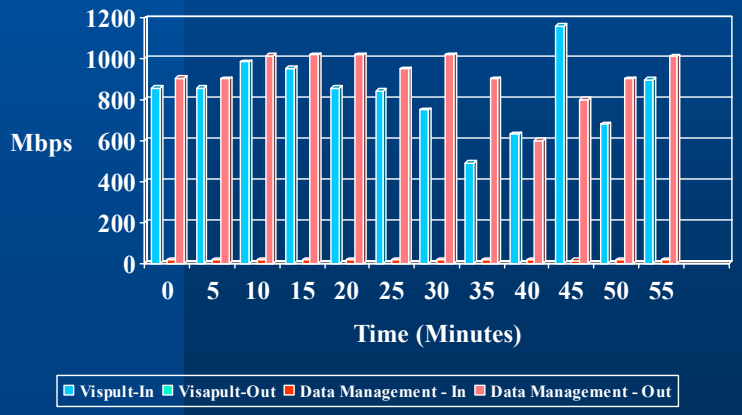
- **Visapult - Using High-Speed WANs and Network Data Caches to Enable Remote and Distributed Visualization**
- **QOS – Enabled Audio Teleportation**
- **A Data Management Infrastructure for Climate Modeling Research**
- **NRL HDTV Video Application**

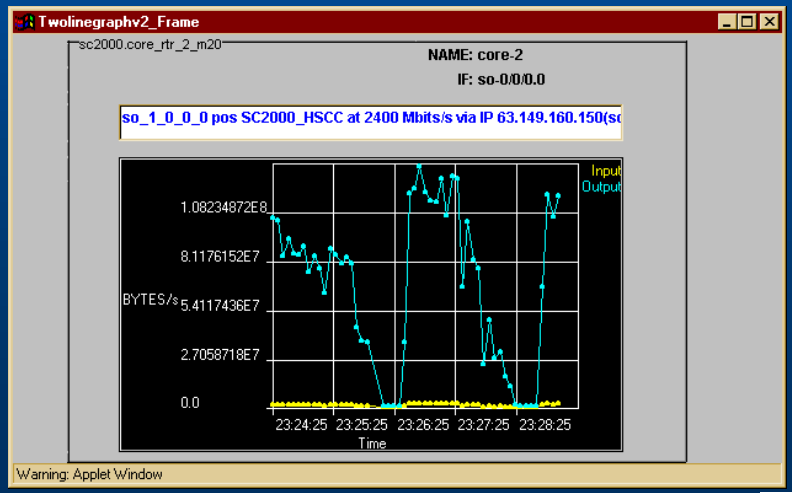


Application Performance

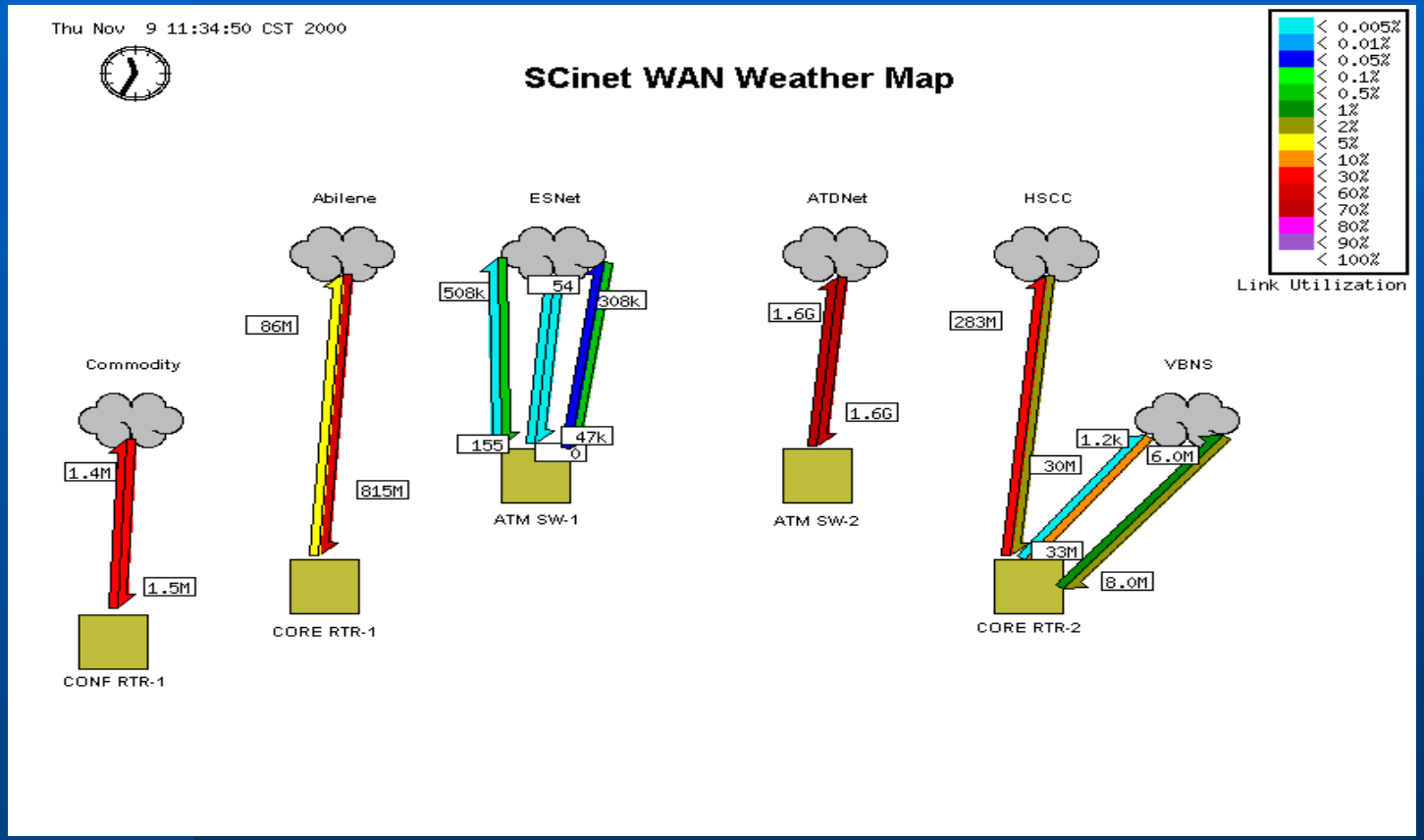


Bandwidth Demonstration of Vispault Application
Application reported 1.48 Gbps
5 second measurements reported 1.56 Gbps
.1 second measurement reported 1.72 Gbps
262 GB transferred in 1 hour (582 Mbps sustained)



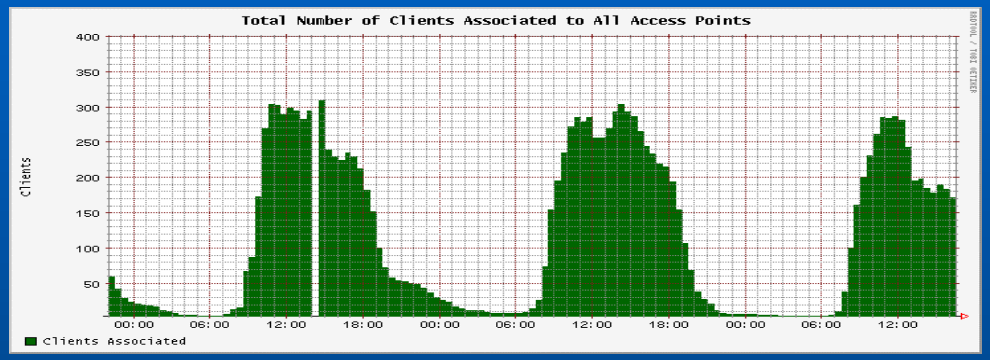


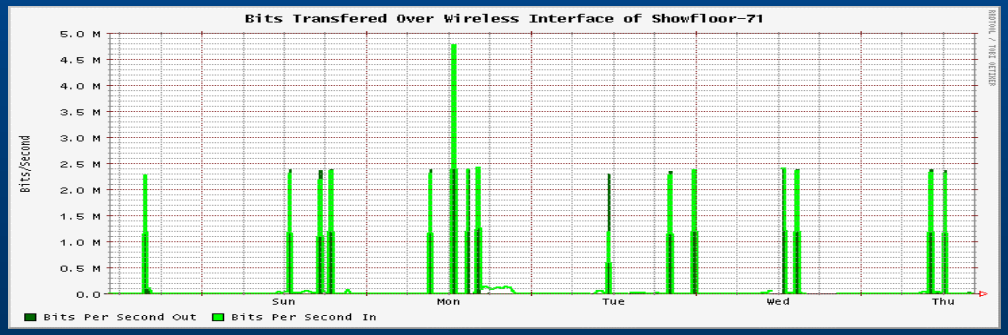
Overall WAN Usage



4.92 Gbps highwater mark

Wireless


Total Number of Clients Associated to All Access Points
Clients
400
350
300
250
200
150
100
50
00:00
06:00
12:00
18:00
00:00
06:00
12:00
18:00
00:00
06:00
12:00
Clients Associated


Bits Transfered Over Wireless Interface of Showfloor-71
Bits/Second
5.0 M
4.5 M
4.0 M
3.5 M
3.0 M
2.5 M
2.0 M
1.5 M
1.0 M
0.5 M
0.0
Sun
Mon
Tue
Wed
Thu
Bits Per Second Out
Bits Per Second In

Summary of SCinet Activity

Category	
Total Miles of Fiber Installed	82.2 hours
Time from first lift to Outside Connectivity	59.7 hours
Miles of Fiber per hour	1.4 fiber miles per hour
Time to first OC-48 Connection (Abilene)	80.2 hours
Total Theoretical Peak External Bandwidth	8.477 Gbps – self limited, 9.477 Gbps actual interface
Estimated Theoretical Peak Show Floor Bandwidth	More Than 130 Gbps
Wireless Coverage Area	Entire Show Area - 207,338 square feet
Total effort of volunteers	Total 75 people = 11.27 people years
Value of volunteer efforts at $200,000 per year	$2,225,000
Estimated Value of donated equipment and services for the tested	Greater than $25,000,000

Observations

- **Xnet – 10 Gigabit Ethernet**
  - Point-to-point network between booths, using Cisco's pre-production 10 Gigabit Ethernet blades for their 6500 series switching routers.
  - CISCO selected the parallel interface to showcase at SC2000.
    - Less than full serialization by intercepting 4 parallel streams and running them out directly as parallel data streams on optical ribbon cable.
    - 20-CPU storage cluster (hooked to the switch through 20 separate Gig-E interfaces) feeding data through a pair of the 10-Gigabit Ethernet cards to a 20-processor compute cluster which was processing the data and rendering images
  - The issues were purely fiber related, or at best ,fiber interface related.
    - Error rate came down to a fully usable level, and the system actually delivered the promised level of performance
    - The 10 GE blades performed well, and did not require and swapping or replacement.
  - The demonstration implementation used "striping" at the application layer and none of the drivers or stacks have been optimized.
  - Eight 1 Gbps streams between a pair of 10 GE ports was consistently demonstrated.
  - Due to fiber limitations it was not possible to utilize all four 10 GE boards.
    - The maximum that would have been transferred was 8 Gbps due to the number of GE feeder ports.

Observations

- **Commodity**
  - **Provided excellent levels of service**
  - **Integrated with Wireless**
- **Wireless –**
  - **It is feasible to provide wireless connectivity to the entire conference**
  - **Interference with rogue devices is very problematic**
    - **Other Access Points**
    - **Other Wireless Devices**
  - **Resolving interference problems is non-trivial**
  - **Load balancing**
    - **Some level feasible with power adjustments**
    - **Not sufficient resolution to provide balanced service**

Observations

- **Production Network**
  - **Intentional network complexity demonstrated interesting aspects of performance and interoperability**
  - **BGP on all the core routers worked well**
  - **Unexpected issue of having HSCC and Commodity traffic both use quest backbone made things complex**
  - **Layer 3 switch performance under real load was interesting to observe (and discrete conclusions about vendor products determined)**
  - **IPv6 support needs improvement – most vendors still do not do it correctly or at all**
  - **Auto configuration of GE does not work in general**

Observations

- **Monitoring and Measurement –**
  - **Bro works very well and can accommodate Gigabit speeds**
  - **Spirent/Adtech provided excellent monitoring at interfaces**
    - **Network complexity required vendor development staff to make a number of modifications/improvements**
    - **Wide area measurement coordination feasible**
  - **SNMP polling is useful in a number of ways**
  - **Detailed monitoring data is significant and complex to evaluate**

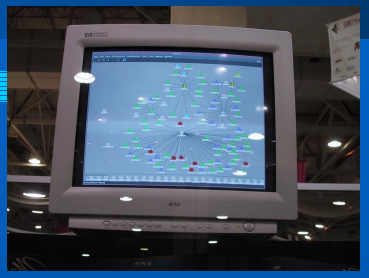
Summary

- It is feasible to build complex networks in short time-frames
  - WAN connectivity is a major factor for success
  - Planning and coordination are mandatory for success
  - Expertise is essential
- Details Monitoring and Measurement are feasible at the 10's of Gbp/s range
  - Application specific
  - System wide

Summary

- The insight gained at unique events like SC is extremely valuable
  - Provide a test bed for other network researchers
  - Provides a gathering place for many vendors to try out inter operation of advanced equipment
- Real Scientific Applications are able to fully utilize bandwidth over the wide area
- It is a lot of work to provide high levels of service in a flexible, technology aggressive environment
  - Can support 1,000 of users, 100's of projects and 10's of intensive demonstrations - but it is not easy


STEPHENS
INSTRUMENTS
JLG LIFT




10/31/2000